An Application of Distributional Semantics for the Analysis of the Holy Quran

Giulia Benotto, Emiliano Giovannetti, Ouafae Nahli
Istituto di Linguistica Computazionale "Antonio Zampolli",
Via Giuseppe Moruzzi 1, 56124 Pisa, Italy
Email: name.surname@ilc.cnr.it

Abstract—**In this contribution we illustrate the methodology and the results of an experiment we conducted by applying Distributional Semantics Models to the analysis of the Holy Quran. Our aim was to gather information on the potential differences in meanings that the same words might take on when used in Modern Standard Arabic w.r.t. their usage in the Quran. To do so we used the Penn Arabic Treebank as a contrastive corpus.**

Index Terms—**Distributional Semantics, the Holy Quran, Classical Arabic, Modern Standard Arabic, Contrastive Linguistics**

I. INTRODUCTION

The computational analysis of literary texts has been one of the first applications of computer science to linguistics. Since the pioneering works of the first computational linguists who, back in the fifties, started to use computers to manipulate texts and language [10], what we now call digital humanities is a relentlessly growing discipline counting many applications in the analysis of literature. In this work, we describe a computational linguistics approach to the analysis of one of the most known, read and studied books: the Holy Quran. In particular, we were interested in verifying if the application of algorithms based on Distributional Semantics could highlight peculiar characteristics about the usage of words inside the Quran. To do so, we applied a contrastive technique by comparing the related terms extracted from the Quran with the related terms extracted from the Penn Arabic Treebank (ATB), composed of texts written in Modern Standard Arabic (MSA). Our attention focused to those pairs found in both corpora but having a significantly different semantic distribution. As it will be described further, we obtained interesting results, showing how the distribution of words can reveal the different semantic nuances that the same words denotes in each corpus.

A review of the scientific literature documenting works similar to ours is reported in Section II. Section III describes the experiment we have conducted, including an introduction to Distributional Semantics and the analysed corpora. The results of the experiment are described and commented in Section IV. Finally, some conclusions are given in Section V.

II. RELATED WORK

In literature, there are many works that apply a computational linguistics approach to the study of humanistic texts -

among which the Holy Quran - with different purposes. This review gathers the works we believe more pertinent w.r.t. ours, and we group them in three categories.

The first one is related to the works aimed at the analysis of literary texts that used the Quran as reference text. Among those, [1] proposes a stylometric model for the study of Quranic oath expressions using application-specific features. Also in [2] the authors focus on the domain of oath statement in the Quranic text, by introducing a stylometric features selection approach evaluated by machine learning algorithms, to find the finest of the features and to study the impact of the features selection on the classifiers performance. Furthermore, there are a few works investigating the authorship of the Quran in which the task of Authorship Attribution has been performed through author discrimination, a method consisting in checking whether two texts are written by the same author or not. In [18], the authors try to make an author discrimination between the Quran and the Hadith (statements said by the prophet Muhammad). This has been performed in different ways: in [19] the Quran and the Hadith are analyzed in a global form and in [20], which represents the continuation of [19], they have been analyzed in a segmental form by extracting 14 text segments from the Quran book and 11 text segments from the Bukhari Hadith. Finally, in [21], the same task has been performed employing two visual analytics clustering methods. However, the purpose of our work is to analyse the Quran from a linguistic point of view by focusing on the semantic distribution of words, rather than the identification of an author or the stylistic analysis of particular expressions.

The second category of works concerns the application of Distributional Semantics (DS) to the analysis of the Quran. [3], [5] and [6] report the construction of a very large corpus of Classical Arabic to be used as a base to study distributional lexical semantics of the Quran and Classical Arabic. Indeed, even if many distributional semantic models are available in the literature, none of them have been applied to the Quran nor to Classical Arabic in general, due to a lack of resources. In [3] authors report the results of two empirical studies; the first is applying a number of probabilistic distributional semantic models to automatically identify lexical collocations in the Quran and the other is applying those same models on the Classical Arabic corpus in an attempt to test their ability of capturing lexical collocations and co-occurrences for a number of the corpus words. The resource has been used to test a

374

new association measure for detecting syntagmatic relations between words, as reported in [4], in which an experimental study to evaluate the performance of the proposed measure is presented. However, though our work is based on the application of DS to the analysis of the Quran, it is not our aim to build up resources for Classical Arabic.

More in general, works describing the application of Distributional Semantics to the analysis of literary texts constitute the third category we take into consideration. In such works, DS is typically used as a theoretical basis in order to verify some hypotheses on specific semantic characteristics of poetries. In more details, in [9] the authors investigated through DS the influence of Lord Byron's work on Thomas Moore trying to find a shared vocabulary or specific formal textual characteristics. In [11] it is argued how distributionalism can support the notion that the meaning of poetry comes from the meaning of ordinary language and how distributional representations can model the link between ordinary and poetic language.

The purpose of our work is to identify potential differences in meanings, or different semantic hints, in the way the same words are used in the Quran w.r.t. how they are used in MSA. The investigation regarding these semantic differences has been conducted by analyzing the differences in their distribution, both in the Quran and in MSA, in accordance with the Distributional Semantics Hypothesis. To the best of our knowledge, this is the first attempt to apply a DS-based algorithm to the semantic analysis of the words used in the Quran.

III. Experiment Description

We performed our experiment using two corpora, the Quran and the Penn Arabic Treebank. Distributional Semantics Models (DSMs) have been applied to both corpora, in order to gather information on the potential differences in meanings that the same words might take on when used in the ATB w.r.t. their usage in the Quran. In the following, subsection III-A gives a brief background on Distributional Semantics, subsection III-B presents the used corpora and, finally, subsection III-C describes the experiment in detail.

A. Distributional Semantics

Distributional semantics (DS) is a theory that pertains to the linguistic tradition, since its history actually began in the context of Zellig Harris' proposal [14] of distributional analysis as the bedrock of linguistics as a scientific discipline. Harris' distributional proposal was conceived for phonemic analysis and later turned into a general methodology. Nowadays DS is used in different linguistic tasks, such as dialectometry, lexico-semantic relation analysis, text mining and paraphrasing.

The assumption behind all DS models is that the notion of semantic similarity can be defined in terms of linguistic distributions [16]. The distribution of an element can then be inferred from the sum of all its contexts, where a context is the setting of a word among the surrounding words i.e., the co-occurrency window. One way to collect this information is to provide, for each word, a list of the co-occurrences of the word and the number of times they have co-occurred.

Schütze ([22],[23]) described the following approach, that has later become standard practice for word-space algorithms: data is collected in a matrix of co-occurrence counts (M), called a co-occurrence matrix; a cell m_{ij} of the co-occurrence matrix records the frequency of occurrence of word i in the context of word j. Such data is then used to build n-dimensional context vectors, defined as the rows or columns of the matrix, which allow us to define (distributional) similarity between words in terms of vector similarity. A convenient way to compute the similarity between the vectors representing such words, is to calculate the cosine of the angle between the two vectors $\vec{w_i}$ and $\vec{w_j}$ representing the words.

B. Corpora

The Quran is the divine religious book of Islam. For Muslims, it contains the actual words of Allah (The God), and it represents the last scripture from Him to humankind. It was revealed to Prophet Muhammad 14 centuries ago in pure Classical Arabic language consisting of 77,430 words [13], which are organized in 114 chapters (*suwar*) of varying lengths [24] that are in turn partitioned into a number of verses (*ʾāyāt*). The Quranic text is an eloquent religious text in a literary style that structurally resembles poetry [24].

In order to perform our experiment, we used the **Quranic Arabic Corpus**[1], an annotated linguistic resource which shows the Arabic grammar, syntax and morphology for each word in the Quran including both roots and lemmas. For verbs, only the root is indicated, while nouns, proper nouns and adjectives have both a root and a lemma. Other parts of speech, such as particles, have only lemmas. The morphological features used to group similar words together make use of the extended Buckwalter transliteration [12].

As a contrastive corpus, we used the **Penn Arabic Treebank** [17], which targets the description of a written MSA corpus from the Agence France Presse (AFP) newswire archives for July-November 2000. This corpus includes 734 stories representing 145,386 words (166,068 tokens after clitic segmentation in the Treebank; the number of Arabic tokens is 123,796). All the corpus is annotated for morphological information, part-of-speech, English gloss, and syntactic structure.

C. Definition of the Experiment

Since its being a statistical method, DS needs huge amount of data in order to provide significant results. So, these techniques are usually applied on large text corpora. For this reason, it has been widely used with news corpora and corpora of common language (such as those derived from Wikipedia, or the WaCky family[2]). This is also the reason why resources like the one described in [3], [5] and [6] have been built for the

---

[1]http://corpus.quran.com/ – Last Accessed: July 28, 2016

[2]ukWaC, deWaC and itWaC are three very large corpora of English, German, and Italian built by web crawling. The corpora contain more than a billion words each, and are thus among the largest resources for the respective languages, as explained in [7].

distributional analysis of Ancient Arabic. However, we believe DS could reveal itself useful for analyzing the style of literary works, despite the fact that their size could be not statistically relevant. Indeed, DS capability of highlighting related terms could show peculiarities of literary texts such as how the style and themes of an author evolve in time, how authors might use some lexical patterns or how such lexical patterns might be used differently in different texts.

To be more precise, the aim of this work was to apply a contrastive technique to isolate those words that have different meanings in the Quran than in the ATB, or those words that assumed different semantic hints in the Quran w.r.t. ATB. We assumed that, if a pair of words has a significantly different semantic distribution in the Quran w.r.t. the distribution the same pair has in the ATB, that would be a clear indication of a different semantic use, in the two corpora, of both the words composing the pair or of one of them.
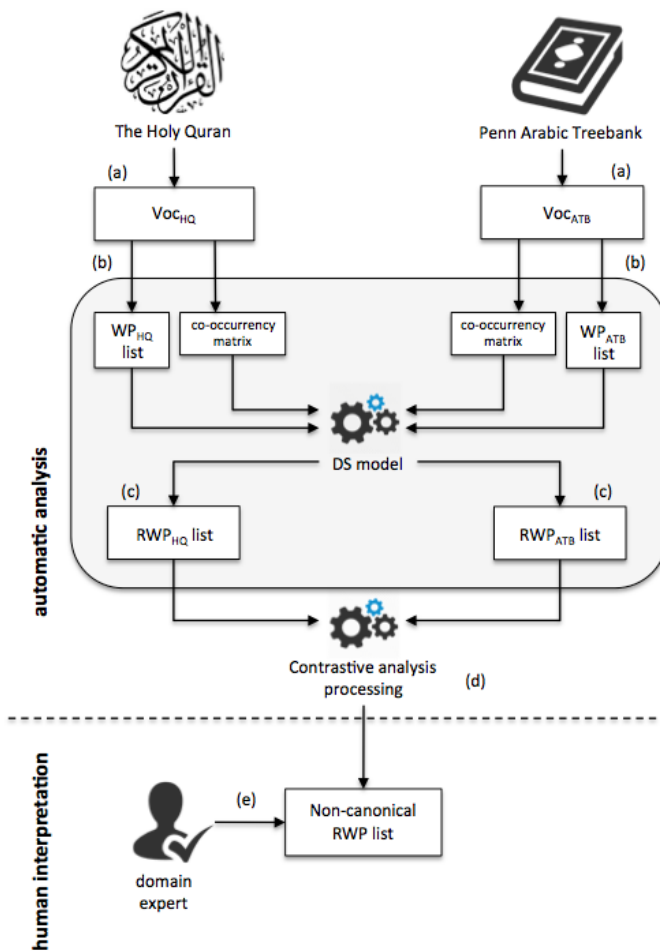

Fig. 1. The steps of the approach used in our experiment.

Figure 1 describes the steps of our experiments. Starting from pos-tagged and lemmatized resources, we built vocabularies for both the Quran and the ATB, that would account for names, verbs and adjectives alongside with their frequencies. We selected names and adjectives in their lemma form while

verbs were selected in their root form since they have no lemma form in the Quranic Corpus (Figure 1(a)). Using the obtained vocabularies, we built two words-by-words matrixes of co-occurrence counts (co-occurrence matrixes) one for the Quran and one for the ATB. Since our corpora were relatively small, and we did not have particular computability issues, we chose not to apply decomposition techniques to reduce the size of the matrixes, consequently, we did not lose any information. We performed different empiric setups of the window size - setting it to 2, 3, 4 and 5. For both corpora and each window size, we built a list of word pairs (as lemmas), where each word of the vocabulary was paired with all the other words with the same PoS of the same vocabulary (Figure 1(b)). The obtained lists and the relative matrixes were used to calculate the cosine between the vector representing each word and the vector representing every other word of the vocabulary [8]. This way, we wanted to evaluate the semantic similarity between the words, by assessing their proximity in the distributional space. Indeed, the assumption underlying the distributional hypothesis [14] states that the semantic similarity of the corresponding words increases as the vectors get closer in the distributional space. Evaluating the cosine of the angle between two vectors, we were then able to assess their relatedness: the more the cosine value tended to 1, the more the two words of the pair under examination were considered to be related.

We then obtained two raw related word pair (RWP) lists, one for the Quran: RWP_{hq} and one for the ATB: RWP_{atb} (Figure 1(c)). We then selected a cosine threshold. All pairs of words having a cosine value above the selected threshold were considered to be somewhat "related". As already stated, we wanted to apply a contrastive technique to isolate those RWP that have different semantic distribution in the Quran than they have in the ATB. In order to find those pairs having a non-canonical distribution, we searched for all RWP_{hq} that present a cosine measure significantly different from the relative ones calculated starting from the ATB (RWP_{atb}). Moreover, we restricted our analysis to the lists relative to the 3 words-sized window, since a manual validation of the obtained RWPs showed them to be the most balanced ones, in accord with what is stated by Kruszewski and Baroni [15]. At the end of this process, we had two lists having the same word pairs with different cosine measures. Since the Quran Corpus and the ATB had two slightly different transliteration, we had to perform a manual alignment of the word pairs contained in the two RWPs lists in order to make them comparable. Indeed, the transliteration used in the ATB, strictly follows Buckwalker's Table[3]. The text of the Quran, instead, contains additional marks at the beginning and end of certain verses, aimed at giving information on the structure of text and recitation, together with pause marks aimed at aiding recitation. Pause marks are rendered in superscript in the middle of verses, at the end of the word they refer to. We then had to get rid of all additional marks from word pairs extracted from the Quran in

---

[3]http://www.qamus.org/transliteration.htm – Last Accessed: July 28, 2016

order to align them to word pairs extracted from the ATB.

Indeed, the final step of our experiment was to compare the lists (Figure 1(d)). For each pair, we calculated the absolute value of the difference between the respective cosine measures (Cos_{diff}) and then sorted the pairs in descending order of such a value. In Table I, we show a list of the first fifteen word pairs that, in the Quran, appear to be non-canonically distributed, i.e. they show a significantly different distribution w.r.t. the ATB.

TABLE I
WORDS HAVING NON-CANONICAL DISTRIBUTION IN THE QURAN

Arabic Words Pair	Transliterated Words Pair	Cos_{diff}
خَطَأ ـ فَتْح	ḫaṭaʾ , taḥrīr	0.999
شُورَى ـ تَحْرِير	fatḥ , šūrā	0.999
عَشِيَة ـ حَسَن	ḥasan , ʿašiyah	0.998
مَنَعَ ـ أَعْطَى	manaʿa , ʾaʿṭā	0.998
حَسَن ـ جَائِر	jāʾir , ḥasan	0.997
شَرِيعَة ـ وَزِير	wazīr , šarīʿap	0.996
اِلْتَقَى ـ قَتَل	qātala , ʾiltaqā	0.995
أَوْرَدَ ـ فَصَل	faṢala , ʾawrada	0.995
جِهَاد ـ وَسِيلَة	wasīlah , jihād	0.994
أَكْثَر ـ فَرِيق	farīq , ʾaktar	0.994
بَيَان ـ دَفْع	dafʿ , bayān	0.993
اِخْتَار ـ ذَهَبَ	dahaba , ʾiḫtāra	0.993
تَوْصِيَة ـ وَزِير	wazīr , tawṣiyah	0.992
مَزِيد ـ سَلِيم	salīm , mazīd	0.991
حَسَن ـ عَامِل	ʿāmil , ḥasan	0.991
أُخْت ـ بَيْت	bayt , ʾuḫt	0.990

IV. ANALYSIS OF THE RESULTS

Overall, our system extracted 160,157 RWP from the Quran and 111,894 from the Penn Treebank. The contrastive analysis we performed showed that 863 RWP could be considered as having significant distribution differences between the Quran and the ATB. Out of these, 9 pairs were more strongly associated in the ATB than they were in the Quran and, conversely, 854 were more strongly associated in the Quran than in the ATB. Table II shows how these RWP are partitioned according to their Cos_{diff} value. An Arabic native speaker manually evaluated the obtained RWPs and suggested to focus on those that are found in the first interval, that is those that present a higher degree of distributional diversity between the Quran and the ATB.

TABLE II
NUMBER OF RWP PARTITIONED ACCORDING TO THEIR Cos_{diff} VALUE

Number of RWPs	Cos_{diff} Interval
470	$0.75 \leq Cos_{diff} < 1$
276	$0.5 \leq Cos_{diff} < 0.75$
126	$Cos_{diff} < 0.5$

In order to deepen our analysis and confirm the relevance of the obtained results, we here provide two examples showing the potential differences in usage and/or meaning between the two corpora.

The first pair of words we analyzed was (*fawoz*, *'iyāb*). This pair of words appears to be more strongly related in the Penn Treebank then in the Quran. We searched for the occurrences of both words in both corpora, in order to understand the different meanings they might hold. The world *fawoz* means "victory", but we found that it assumes different semantic nuances in the Penn Treebank w.r.t. the Quran. In the Penn Treebank, indeed, both *fawoz* and *'iyāb* are often found occurring while talking of sports, namely of athletic matches that are constituted by a first leg and a rematch. *fawoz* then means "victory", as in an occasion when someone wins a game or a competition, but without any further meanings. In the light of this fact, it is clearer why *fawoz* and *'iyāb* have a weaker association strength in the Quran. We were then curious to understand to which words *fawoz* is more related to in the Quran, in order to grasp the semantic meaning this word takes on in this Text. Analyzing all the pairs of words that present a higher degree of association in the Quran w.r.t. the ATB, we isolated three examples, in which the difference in strength association is higher than 0.9 since the words composing these pairs are significantly more associated in the Quran than in the ATB.

خَالِد ـ فَوْز	fawoz – ḫālid	0.966
نَعِيم ـ فَوْز	fawoz – na'īm	0.946
مَشْهَد ـ فَوْز	fawoz – mašhad	0.937

As it appears from Table III, in the Quran *fawoz* (which means attainment/victory) is strongly associated with words like *ḫālid* (which means eternal/everlasting), *na'īm* (which means bliss) and *mašhad* (which means both spectacle and testimony), since they tend to occur in similar contexts. All these words, in the Quran, are strongly related to the concept of the "last day", especially regarding the Victory that believers obtains after Death, when they can finally witness all the beatitudes that reside in the Afterlife:

بُشْرَاكُمُ الْيَوْمَ جَنَّاتٌ تَجْرِي مِن تَحْتِهَا الْأَنْهَارُ خَالِدِينَ فِيهَا ذَلِكَ هُوَ الْفَوْزُ الْعَظِيمُ

"Glad tiding for you this Day! Gardens under which rivers flow (paradise), to dwell therein forever! Truly, this is the great success!" (sūrah 57:12).

فَوَيْلٌ لِّلَّذِينَ كَفَرُوا مِن مَّشْهَدِ يَوْمٍ عَظِيمٍ

"so woe to the disbelievers from the Meeting of a great Day (i.e. that of Resurrection)" (sūrah 19:37).

فَأَثَابَهُمُ اللَّهُ بِمَا قَالُوا جَنَّاتٍ تَجْرِي مِن تَحْتِهَا الْأَنْهَارُ خَالِدِينَ فِيهَا وَذَلِكَ جَزَاء الْمُحْسِنِينَ

"So because of what they said, Allāh rewarded them Gardens (in Paradise), they will abide therein forever. Such is the reward of Al-Musinīn (the good-doers)" (sūrah 5:85).

فَرَوْحٌ وَرَيْحَانٌ وَجَنَّةُ نَعِيمٍ

"(There is for him) Rest and provisions, and a Garden of Delights (Paradise)" (sūrah 56:89).

So *fawoz*, which obviously still means "victory", assumes a different nuance of meaning in the Quran than that it commonly assumes in the ATB, strongly and completely correlated with the themes treated in the Holy Text.

سُور ـ مِصْر	miṣr – sūr	0.955
عُقْب ـ مِصْر	miṣr – 'uqb	0.846
حَاشِيَة ـ مِصْر	miṣr – ḫašyah	0.805

Another interesting example of the peculiarities that can be highlighted with this kind of contrastive analysis, is provided by the distributional behaviour of the word *miṣr*. In the Penn Treebank *miṣr* is only used according to the meaning it has in MSA, i.e. the geographic area commonly known as Egypt.

In the Quran, Table IV shows that the word *miṣr* is strongly semantically related to the words *sūr* (which means "wall"), as well as with *'uqb* (which means "the final end"), and *ḫašyah* ("fear"), while these semantic associations are much less strong in the ATB.

Deepening the analysis, we searched for the meaning of *miṣr* in the *'al-mu'ǧam al-wasīṭ* dictionary[4], where we found that, indeed, miṣr has another sense (as border) in Classical Arabic:

المِصْرُ : الحاجِزُ بين الشيئَين ، أو بين الأَرضين .

Al-miṣru : the border between two things or two properties/domains

and

و (مِصْرُ) : هي البلدُ العربيُّ المعروف .

(*miṣru*): the well known Arabic country

This finding may provide a possible explanation of the relatedness between, at least, the word *miṣr* (as "border") and the words *sūr* ("wall") and *'uqb* ("the final end").

Of course, in order to provide an interpretation of these data, the final word is up to a Quranic scholar.

V. CONCLUSION

The application of our contrastive technique to the Quran and the Penn Arabic Treebank proved to be fruitful: the number of what we have considered to be very interesting word pairs (470) confirms that it is worth to apply these techniques for the discovery of words having particular semantic nuances in ancient texts. As a matter of fact, it can be very hard to detect semantic correlations between words appearing in different parts of the same text (i.e. in absentia) without adopting a Distributional Semantics approach. To conclude, we believe that this techniques may provide scholars and linguists with interesting causes for reflection to approach the contrastive study of Classical and Modern Standard Arabic.

[4]'al-mu'ǧam al-wasīṭ dictionary, fourth edition, maktabat al-shuruwq ad-dawliyyah, Egypt, 2004.

REFERENCES

[1] A. Alqurneh, A. Mustapha, M. A. Azmi Murad, and N. M. Sharef, *Stylometric model for detecting oath expressions: A case study for Quranic texts*, Digital Scholarship in the Humanities, Dec. 2014.

[2] A. Alqurneh and A. Mustapha, *The Impact of Oath Writing Style on Stylometric Features and Machine Learning Classifiers*, Journal of Computer Science, vol. 11, no. 2, pp. 325329, Feb. 2015.

[3] M. Alrabiah, N. Alhelewh, A. Al-Salman, and E. S. Atwell, *An empirical study on the Holy Quran based on a large classical Arabic corpus*, International Journal of Computational Linguistics (IJCL), vol. 5, no. 1, pp. 113, 2014.

[4] M. Alrabiah, A. Al-Salman, and E. S. Atwell, *A New Distributional Semantic Model for Classical Arabic*,

[5] M. Alrabiah, A. Al-Salman, and E. S. Atwell, *The design and construction of the 50 million words KSUCCA*, Proceedings of WACL2 Second Workshop on Arabic Corpus Linguistics, pp. 58, 2013.

[6] M. Alrabiah, A. Al-Salman, E. S. Atwell, and N. Alhelewh, *KSUCCA: a key to exploring Arabic historical linguistics*, International Journal of Computational Linguistics (IJCL), vol. 5, no. 2, pp. 2736, 2014.

[7] M. Baroni, S. Bernardini, A. Ferraresi and E. Zanchetta, *The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora*, Language Resources and Evaluation 43 (3): 209-226, 2009.

[8] M. Baroni and A. Lenci, *Distributional memory: A general framework for corpus-based semantics*, Computational Linguistics, 36(4):673721, 2010.

[9] P. Buitelaar, N. Aggarwal, and J. Tonra, *Using Distributional Semantics to Trace Influence and Imitation in Romantic Orientalist Poetry*, in AHA!-Workshop 2014 on Information Discovery in Text, 2014.

[10] R. Busa, *The Use of Punched Cards in Linguistic Analysis*, Punched Cards-Their Application to Science and Industry, 2nd edition, R.S. Casey et al. ed., New York, Reinhold, pp. 357-373, 1958.

[11] A. Herbelot, *The semantics of poetry: A distributional reading*, Digital Scholarship in the Humanities, vol. 30, no. 4, pp. 516531, 2015.

[12] T. Buckwalter, *Buckwalter Arabic transliteration*, http://www.qamus.org/transliteration.htm, 2002.

[13] K. Dukes, and N. Habash, *Morphological annotation of Quranic Arabic*, The seventh international conference on Language Resources and Evaluation (LREC-2010), Valletta, Malta, 2010.

[14] Z. S. Harris, *Distributional structure*, Springer, 1970.

[15] G. Kruszewski, and M Baroni, *Dead parrots make bad pets: Exploring modifier effects in noun phrases*, Lexical and Computational Semantics (* SEM 2014), 2014.

[16] G. A. Miller and W. G.Charles, *Contextual correlates of semantic similarity*, Language and cognitive processes, 6(1):128, 1991.

[17] M. Maamouri, A. Bies, H. Jin and T. Buckwalter, *Arabic Treebank*, Philadelphia: Linguistic Data Consortium, 2003.

[18] H. Sayoud, *Author discrimination between the Holy Quran and Prophets statements*, Literary and Linguistic Computing, vol. 27, no. 4, pp. 427444, Dec. 2012.

[19] H. Sayoud, *Segmental Analysis-Based Authorship Discrimination between the Holy Quran and Prophets Statements*, Digital Studies / Le champ numérique, North America, 0, oct. 2015.

[20] H. Sayoud, *Automatic authorship classification of two ancient books: Quran and Hadith*, 2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA). IEEE, 2014.

[21] H. Sayoud, *A Visual Analytics based Investigation on the Authorship of the Holy Quran*, Proceedings of the 6th International Conference on Information Visualization Theory and Applications (VISIGRAPP 2015), 2015.

[22] H. Schütze, *Dimensions of meaning. Supercomputing92*, Proceedings, pp. 787–796. IEEE, 1992.

[23] H. Schütze, *Word space*, Advances in Neural Information Processing Systems 5. Citeseer, 1993.

[24] A. T. Al-Taani and A. M. Al-Gharaibeh, *Searching Concepts and Keywords in the Holy Quran*, Jordan: Department of Computer Science, Yarmouk University, 2010.

[25] W. Zaghouani, A. Hawwari, and M. Diab, *A pilot propbank annotation for Quranic Arabic*, Proceedings of the First Workshop on Computational Linguistics for Literature (NAACL-HLT2012), 2012.